

Big datasta ja analytiikasta apua tiedolla johtamiseen

Niina Lehmus



Tekijä(t) Niina Lehmus	
Koulutusohjelma Tietojenkäsittelyn koulutusohjelma	
Opinnäytetyön otsikko Big datasta ja analytiikasta apua tiedolla johtamiseen	Sivu- ja liitesivumäärä 30
Opinnäytetyön otsikko englanniksi Big data and analytics to support knowledge management	
<p>Tämän opinnäytetyön aiheena on rakenteellisen ja rakenteettoman tiedon yhdistäminen verkkoanalytiikasta kerätyyn tietoon tiedolla johtamisen tueksi. Työ tehtiin toimeksiantona työnantajalleni Invenco Oy:lle. Tavoitteeksi asetettiin big data ja verkkoanalytiikan ymmärryksen ja osaamisen täydentäminen Invenconssa. Lisäksi tavoitteena työllä oli analytiikkatyökalun kehittämisvaiheiden, ominaisuuksien ja toiminnollisuuksien kuvaaminen.</p> <p>Opinnäytetyön tietoperustassa kuvataan tiedon tasoja, tiedon muotoja ja erilaisia tiedon hyödyntämismahdollisuuksista. Lisäksi kuvataan tiedon jalostamisen viitekehys datasta tiedoksi ja edelleen viisaudeksi. Tietoperusta sisältää esittelyn big datasta sekä verkkoanalytiikasta, joita hyödyntämällä yritykset hakevat vastauksia päätöksen teon tueksi. Lähdemateriaalina teoriaosuudelle käytettiin ammattikirjallisuutta sekä verkosta löydettyjä artikkeleita.</p> <p>Invenco osallistui Digilen ja Tekesin Need For Speed –ohjelmaan. Sen lopputuloksena Invencon syntyi pilottiversio analytiikkatyökalusta. Empiirisessä osassa kuvataan Need For Speed –ohjelman aikana Invencon toteuttama pilotti analytiikkatyökalusta. Pilotti toteutettiin referenssiasiakkaalle, jonka tarpeet ja vaatimukset huomioitiin analytiikka työkalun suunnittelussa ja pilotin toteutuksessa.</p> <p>Opinnäytetyön suunnittelu aloitettiin keväällä 2014 samaan aikaan Need For Speed –ohjelman Invencon työosuuden kanssa. Tässä kohtaa syntyi rakenne opinnäytetyön tietoperustalle. Syksyllä 2015 käynnistettiin analytiikkatyökalun pilottiversion toteutus. Empiirisen osuuden kirjoitus aloitettiin pilotin valmistuttua vuodenvaihteessa 2015.</p>	
Asiasanat Tieto, Big data, Analytiikka, Kävijäseuranta, Mittarit, Tiedolla johtaminen	

Author(s) Niina Lehmus	
Degree programme Business Information Technology	
Report/thesis title Big data and analytics to support knowledge management	Number of pages 30
<p>The subject of this thesis is how to combine structured and unstructured data to the data that have been collected from web analytics to support knowledge management. The work was commissioned by my employer Invenco Oy. The objective was to complement the understanding and knowledge of big data and web analytics in Invenco. In addition, the objective of the research project was to define the development stages, features and functionality of analytics tool.</p> <p>The theory part of the thesis describes the levels of data, forms of data and variety of potential uses of the data. In addition, the data processing frame of how to process data into information and further into wisdom is described. The theory part includes a presentation of big data and web analytics which utilizes companies' search answers to support decision making. The source material used in the theory part was professional literature and online articles found from the Internet.</p> <p>Invenco attended the Need For Speed program which was organized in co-operation with Digile and Tekes. As a result the pilot version of the analytics tool was created to Invenco. In the empirical part is described the pilot of the analytics tool carried out during the Need For Speed program by Invenco. The pilot was carried out for a reference customer, whose needs and requirements were taken into account when designing and implementing the pilot of the analytics tool.</p> <p>Designing of the thesis started in spring 2014 at the same time with Invenco's work in the Need For Speed program. This is where the structure of the theory part of the thesis was created. In the autumn of 2015 the implementation of analytics pilot was launched. Writing the empirical part was initiated after the pilot was completed at year end 2015.</p>	
Keywords Data, Big data, Analytics, Web analytics, Metrics, Knowledge management	

Sisällys

1	Johdanto	1
2	Tiedon näkökulmat	3
2.1	Tiedon tasot	3
2.2	Tiedon lajit	4
2.3	Big data eli massatieto	4
2.4	Big datan lähteet	5
2.5	Big datan hyödyntäminen	6
3	Tiedosta päätöksiin	9
3.1	Tarpeiden kartoittaminen	10
3.2	Tavoitteiden tunnistaminen	11
3.3	Mittareiden määrittäminen	11
3.4	Tiedon keruu	13
3.5	Tiedon analysointi	14
3.6	Parannuksien toteuttaminen	15
4	Case: Big datan ja analytiikan hyödyntäminen	17
4.1	Analytiikkatyökalulle asetut tarpeet ja tavoitteet asiakasnäkökulmasta	18
4.2	Liiketoimintalähtöiset mittarit perustuvat asetettuihin tavoitteisiin	18
4.3	Analysoitava tieto kerätään valituista lähteistä	20
4.4	Mittarit visualisoidaan reaaliaikaisina tai staattisina raportteina	20
4.5	Analytiikkatyökalun toiminnolliset kokonaisuudet	24
4.6	Valittu tekniikka tukee analytiikkatyökalun toiminnallisuutta	24
5	Pohdinta	26
	Lähteet	28

1 Johdanto

Internet on tänä päivänä täynnä yrityksille arvokasta raakatietoa hyödynnettäväksi tiedolla johtamisen ja päätöksenteon tueksi. Nykyaikainen teknologia mahdollistaa kiihtyvällä vauhdilla yhä monimuotoisemman tiedon keräämisen, tallentamisen ja hyödyntämisen. Yrityksien toiminnan kehittäminen, tehostaminen sekä varmistaminen perustuvat olemassa olevaan tietoon yrityksen järjestelmissä, mutta yhä enemmän tietoon, joka kerätään Internetistä. Yrityksen sisäisten järjestelmien rakenteellinen tieto antaa näkemyksen siitä kuinka liiketoiminta toimii juuri nyt. Yhdistämällä rakenteellinen tieto analytiikkatietoon saadaan uusia ulottuvuuksia ja näkymiä liiketoiminnan johtamisen tueksi.

Opinnäytetyö käsittelee verkkoanalytiikan avulla verkosta ja sosiaalisesta mediasta kerätyn raakatiedon sekä yrityksen rakenteellisen tiedon kuten myynti-, talous-, asiakas- ja tuotantotiedon sekä näiden yhdistelmien hyödyntämistä yrityksissä päätöksenteon ja johtamisen tukena. Opinnäytetyön tietoperustan tarkoitus on selvittää perusteet tiedon tasoisista, tiedon muodoista sekä erilaisista tiedon hyödyntämismahdollisuuksista sekä perusperiaatteen eli viitekehyksen tiedon jalostumiselle. Opinnäytetyön tarkoituksena on lisäksi havainnollistaa analytiikan hyödyntämistä päätöksenteon tukena case-esimerkin avulla. Teknologinäkökulma sekä markkinoilla olevien teknologioiden vertailu on rajattu opinnäytetyön ulkopuolelle.

Opinnäytetyön toimeksiantajana on työnantajani Invenco Oy. Invenco keskittyy asiakasyritystensä suorituskyvyn johtamiseen sekä raportoinnin kehittämiseen. Tärkeimmät sovellusosaamisalueet ovat budjetoinnin, ennustamisen ja suunnittelun järjestelmät, tietovarastointi ja tiedon integrointi sekä liiketoimintatiedon hallinta (Business Intelligence). Edellä mainittujen lisäksi Invenco täydentää osaamisalueitaan big data ja analytiikka ratkaisulla.

Invenco osallistui Syksyllä 2014 Digilen ja Tekesin Need For Speed –ohjelmaan. Ohjelman yhtenä työkokonaisuutena oli arvon tuottaminen reaaliaikaisesti yrityksissä. Työkokonaisuudessa haettiin uusia lähestymistapoja, työkaluja, menetelmiä ja toimintamalleja hyödynnettäväksi yrityksissä. Invenco täydensi työkokonaisuudessaan osaamistaan verkkoanalytiikkaosaamisella sekä kerrytti käyttökokemuksia verkkoanalytiikan työkaluista. Työkokonaisuus toteutettiin aivoriihiyppisesti, innovoivasti sekä pilotoiden Invencon liiketoimintatiedon hallintaan erikoistuneiden konsulttien toimesta. Need for Speed –ohjelman Invencon työkokonaisuuden lopputuloksena syntyi yleisluonteinen toimialariippumaton analytiikkatyökalu, jonka pilottiversio toteutettiin Microsoftin Azurella. Työkalu toimii reaaliaikaisena käyttöliittymänä big data –järjestelmille. Sen avulla tietoja kerätään sosiaalisesta mediasta, julkisilta sivustoilta ja yhdistetään yrityksen rakenteelliseen tietoon sekä muoka-

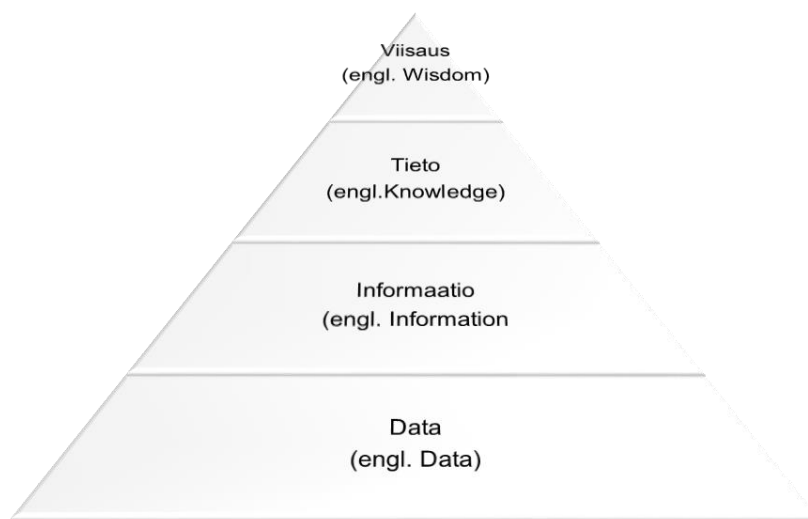
taan esitettävään muotoon raportoitavaksi ja analysoitavaksi käyttäjäorganisaatiolle. Opinnäytetyön empiirisessä osuudessa esitellään Invencon työkokonaisuudesta syntyneen analytiikkatyökalun ominaisuuksia sekä toiminnallisuutta.

2 Tiedon näkökulmat

Antiikin Kreikan filosofin Platonin määritelmä tiedolle oli, että väite muuttuu tiedoksi, jos se on tosi ja perusteltu. ”Käsitys on tietoa silloin kuin sen puolesta pystyy esittämään jonkun oikeutuksen, selityksen hyvän perustelun tai määritelmän” (Lukion filosofia 2001).

2.1 Tiedon tasot

Data, informaatio, tieto ja viisaus ovat tiedon tasoja, joiden rakenteellista ja toiminnallista suhdetta kuvataan yleisesti pyramidina (kuvio 1). Data jalostetaan informaatioksi. Informaatio täydentyy tiedoksi. Tietoa täydennetään kokemuksella ja tietämyksellä, jolloin syntyy viisaus. (Tietojohdamisen tutkimuskeskus NOVI 2014.)



Kuvio 1. Tiedon tasot kuvattuna pyramidina (Mohanty & Jagadeesh & Srivatsa 2013, 2)

Data (engl. Data) on numeroita, merkkejä tai symboleja, yksittäisiä tiedon jyvää, joita myös raakadataksi kutsutaan. Raakadataa ovat esimerkiksi säähavainnot, pankki- tai liittymätahtumat, kurssiarvosanat, urheilusuoritukset tai vaikkapa tuotehinnat. Se on potentiaalista informaatiota, joka ei yksinään ole kovin merkityksellistä. Hyödylliseksi informaatioksi (engl. Information) data muuttuu kuin sitä, yhdistellään, käsitellään, jalostetaan ja tulkitaan. Informaatio on yleisesti myös välitettävää, siirrettävää tai viestitettävää tietoa. Kuvaaja päivälämpötiloista tai urheilusuorituksista, tili- ja opintorekisteriote sekä keskihinnat ja niiden muutokset ovat esimerkkejä informaatiosta, jossa on sisältöä ja jota voidaan hyödyntää viestinnän apuna. Informaatio muuttuu edelleen tiedoksi (engl. Knowledge) kun se täydentyy kokemuksella, opiskelun tai ajattelun kautta. Tiedon syntymiseen vaikuttaa myös aikaisempi tietämys asiasta. Viisauden (engl. Wisdom) katsotaan olevan korkeata-

soisempaa tietoa, joka syntyy tiedon perustasoista: data, informaatio ja tieto. Viisaus on kyky käyttää tietoa omassa toiminnassaan. Se on tulkittua informaatiota, joka on syntynyt toiminnasta saadun kokemuksen, osaamisen sekä tiedon ja ymmärryksen kautta. Usein viisautta kutsutaan myös hiljaiseksi tiedoksi ja sosiaaliseksi pääomaksi, koska sen perustana ovat organisaation tai yksilöiden osaaminen, toiminta tai kokemukset. Tiedolla johtamisen edellytys on pyramidin ylimmän tason eli viisauden saavuttaminen. (Tietojohtamisen tutkimuskeskus NOVI 2014.)

2.2 Tiedon lajit

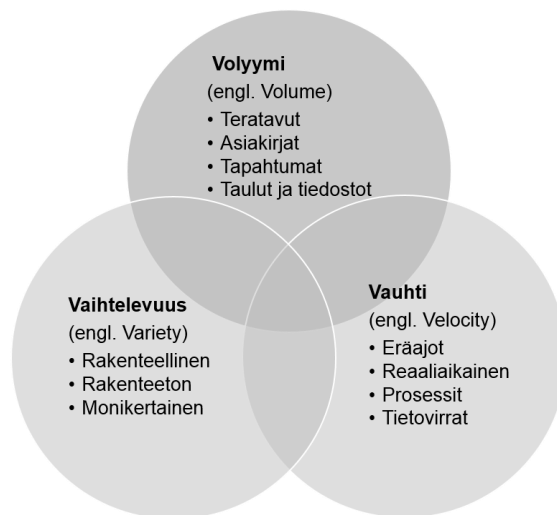
Tieto on henkilökohtaista tai julkista, jota myös avoimeksi tiedoksi kutsutaan. Tieto on rakenteellista tai rakenteetonta. Se voi olla tietoa tiedosta eli metatietoa, joka tarkoittaa varsinaista tietoa lisäämällä tietoon esimerkiksi luontiajankohdan, tiedostotyyppin sekä luojan ja muokkaajan. Tieto voi olla yksittäistä tai se voi muodostaa jatkuvasti lisääntyviä valtaisia tietomassoja. Julkinen tieto on yleisesti sähköisessä muodossa olevaa sisältöä, joka on Internetissä vapaasti käytettävissä, muokattavissa ja jaettavissa eteenpäin. Esimerkkinä julkisesta tiedosta käytetään yleisesti Wikipedia-tietosanakirjaa. Tieto, joka on tallennettu tietokannan kenttään tai tiedostoon on rakenteellista tietoa. Rakenteellista tietoa kutsutaan myös strukturoiduksi ja jäsennellyksi tiedoksi. Sille määritellään pituus ja se voi olla numeroita, kirjaimia tai päivämääriä. Tieto tallennetaan tietokantaan sille varattuun paikkaan ja sillä on yleensä riippuvuuksia toisiin tietoihin. Rakenteellinen tieto on nopeasti ja helposti saatavilla, tallennettavissa sekä analysoitavissa. Rakenteellista tietoa ovat esimerkiksi yrityksen tuote- ja asiakastiedot tai yrityksen myynti- ja liikevaihtotiedot. Tietovarastot ovat rakenteellista tietoa. Tietovarastoihin, jotka yleisesti ovat relaatiotietokantoja, kerätään, yhdenmukaistetaan ja yhdistetään dataa yrityksen operatiivisista järjestelmistä. Tietovarastot tukevat yrityksessä tuotettavaa raportointia ja datan analysointia. Rakenteellista tietoa käsitellään hyödyntäen tietokantoja varten kehitettyä standardisoitua kyselykieltä (SQL, Structured Query Language). Rakenteeton tieto on vastakohtaisesti järjestelmätöntä tietoa kuten tekstiä tai esimerkiksi multimediaa: videoita, kuvia ja ääntä. Rakenteetoman tiedon määrän oletetaan kasvavan merkittävästi sekä huomattavasti nopeammin kuin rakenteellisen tiedon määrän. Big data on esimerkki rakenteettomasta tiedosta, jota kertyy nopeasti valtaisia tietomassoja. (IT Business Edge 2015a; IT Business Edge 2015b.)

2.3 Big data eli massatieto

”Big data tarkoittaa käsittämättömän suurten ja järjestämättömien tietomassojen keräämistä, säilyttämistä ja ennen kaikkea analysointia tietoteknisten ratkaisujen avulla” (Vakuri 20.6.2013). Big data, joka suomeksi kääntyy massatiedoksi tai isoksi tiedoksi, on ollut

jo useita vuosia yleiskäsite suurille tietomäärille. Koska suomenkielinen termistö aiheen ympärillä ei ole kovinkaan vakiintunutta, käytän opinnäytetyössäni englanninkielistä käsitettä ”big data”. Käsitteenä big datan merkitys on hyvin erilainen eri toimijoille. Osa korostaa datan suurta määrää, osa sen monimuotoisuutta, osa datavirran nopeutta ja reaaliaikaisuutta. Osalle toimijoista merkityksellisintä ovat tarvittavat tekniikat ja analyyseista saatavat hyödyt. (Liikenne- ja viestintävirasto 2014, 7.)

Big dataa kuvataan yleensä kolmella v-kirjaimella alkavalla sanalla: volyymi, variaatio, vauhti (kuvio 2). Volyymi tarkoittaa suuria datamääriä, joita syntyy jatkuvasti. Erilaiset tapahtumatiedot ja esimerkiksi verkkoanalytiikasta saatu klikkaustieto ovat hyviä esimerkkejä suurista datamääristä. Variaatio tarkoittaa erilaista, vaihtelevaa ja monimuotoista dataa. Se voi olla rakenteellista tai rakenteetonta. Se voi sisältää esimerkiksi kuvia, ääntä, videoita tai tekstiä ja sitä syntyy muun muassa sähköposteista ja sosiaalisen median kanavissa. (Krishnan 2013.) Vauhti kuvaa datan nopeaa syntymistä reaaliajassa: datan syöttöä ja käyttöä. Hyvänä esimerkkinä nopeudesta ovat erilaiset analyysit sosiaalisen median tiedoista tai erilaisten tietojen yhdistäminen hyödyntäen useita tietolähteitä. (Prajapati 2013.) Kolmen yleisimmin käytetyn v-määritelmän lisäksi big dataa kuvataan kahdella muulla v-kirjaimella alkavalla sanalla. Value, joka kuvastaa datan arvoa liiketoiminnalle sekä veracity, joka tarkoittaa datan ymmärrettävyyttä ja laatua.



Kuvio 2 Big datan kolme ulottuvuutta (Cheesman 2014)

2.4 Big datan lähteet

Raakadataa muodostuu yhä enemmän ja enemmän lähes kaikista verkkoon liitetystä laitteista ja järjestelmistä. Kansainvälisen tutkimusyhtiön International Data Corporation (IDC) arvion mukaan erilaisten laitteiden kuten tablettien ja älypuhelimien, palveluiden ja

anturijärjestelmien yleistymisen myötä digitaalinen avaruus kasvaa vuosittain 40 %. Vuonna 2020 sen koko tutkimusyhtiön mukaan on arviolta 44 biljoonaa gigatavua. (Turner, Gantz, Reinsel & Minton 2014.)

Big datan lähteitä ovat erilaiset verkkoon kytketyt laitteet ja järjestelmät, jotka lähettävät monimuotoista dataa tallennettavaksi tietovarastoihin. Data voi olla esimerkiksi aika- ja paikkatietoja anturijärjestelmistä tai mittaustietoa laitteistojen toiminnasta. Se voi olla verkkosivujen klikkaustietoja, sovellusten lokitietoja, asiakaspalautteita tietojärjestelmistä tai sosiaalisesta mediasta saatua reaaliaikaista tietoa kuten videoita ja kuvia. (Krishnan 2013.) Krishnan (2013) luokittelee raakadatan lähteiden mukaisesti: kone- ja laitetiedot, järjestelmien lokitiedot, verkkoanalytiikka, sähköpostit, sosiaalinen media, paikkatiedot sekä matkapuhelinverkot. Alla olevassa taulukossa (taulukko 1) on esitetty yhteenvetona Krishnan (2013) listaamat lähteet ja esimerkkejä raakadatasta, jota lähteistä kerätään.

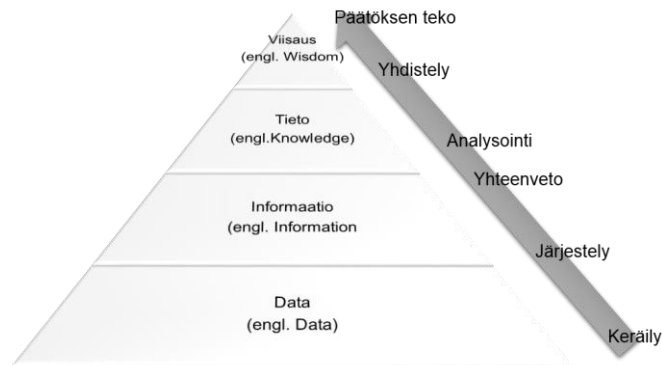
Taulukko 1. Esimerkkejä lähteistä ja niiden tuottamista tiedosta (Krishnan 2013)

	Esimerkkejä lähteistä	Esimerkkejä lähteen tuottamasta tiedosta
Konetiedot ja järjestelmien lokitiedot	Anturit Henkilöskannerit Matkapuhelimet ja muut Mobiililaitteet Moottorit Radiosignaalit Robotit Röntgenlaitteet Sateliitit Tietokoneet(autot, lentokoneet, junat jne.)	Käyttäjätietoa Paikkatietoa Tapahtumatietoa Toiminnon ajankohtatiedot
Matkapuhelimet	Matkapuhelinverkot	Kuvia Musiikkia Videoita
Paikkatiedot	Älypuhelinsovellus, GPS (Global Positioning System) Puhelimet Kamerat Videot	Ajankohta Sijainti
Sosiaalinen media	Sosiaalisen median kanavat kuten Facebook, Twitter, Instagram, LinkedIn	Keskustelut Näkemykset Tykkäämiset
Sähköpostit	Sähköpostijärjestelmät	Sähköposteja liitetietoineen
Verkkoanalytiikka	Verkkosivut	Klikkauspolut Vierailijat ja Vierailut Vierailijoiden käyttämät hakusanat Vierailijoiden tulotapa sivustolle Suosituimmat sivut

2.5 Big datan hyödyntäminen

Määritelmänsä mukaan big data ei ole vain valtavien tietomassojen keräämistä ja tallentamista. Se on myös datan jalostamista informaatiosta tiedoksi ja tietämykseksi sekä vii-

saudeksi (Big Data 2015). Vasta datan järjestely, yhdistely ja analysointi tekevät kerätystä tietomassasta merkityksellistä ja arvokasta liiketoiminnalle ja sitä voidaan hyödyntää tiedolla johtamisessa ja päätöksen teon tukena (kuvio 3).



Kuvio 3. Datasta päätöksenteon tueksi (Mohanty&Jagadeesh&Srivatsa 2013, 2)

Big data tuottaa lisäarvoa yrityksille riippumatta toimialasta tai yrityksen koosta. Finanssi-toimialalla dataa käytetään petoksien havaitsemiseksi ja etsitään mahdollisuuksia onnis-tuneisiin kauppoihin. Logistiikka hyödyntää dataa kuljetusreittien optimoinnissa hakien kustannuksien säästöjä sekä toiminnan tehostamista. Media- ja telekommunikaatiotoi-mialalla big dataa hyödynnetään käyttäjien mielenkiinnon kohteiden sekä käyttäytymisen selvittämiseen. Tavoitteena on tarjota asiakkaille parempia palveluita ja parantaa asiakas-tyytyväisyyttä. Terveystieteiden toimialalla dataa kerätään optimaalisen hoidon tarjoa-miseksi potilaille ja kaupan toimialalla etsitään vastauksia kuluttajien käyttäytymiseen ja tarjotaan räätälöityjä, kohdennettuja kampanjoita esimerkiksi sosiaalista mediaa hyödyn-täen. (Mohanty&Jagadeesh&Srivatsa 2013, 9-10.)

Suurten tietomäärien keräämisen ja tallentamisen ovat mahdollistaneet alentuneet tallen-nus- ja prosessointikustannukset ja laskentamenetelmien kehittyminen. Kehittyneet tieto- ja viestintäteknologiat mahdollistavat lisäksi tiedon reaaliaikaisen ja automatisoidun käsit-telyn. (Liikenne- ja viestintäministeriö 2014, 8.) Big dataa hyödynnetään Suomessa jo lä-hes kaikilla toimialoilla: liiketoiminnan raportoinnissa, myynnin ja markkinoinnin tehostami- sessa ja esimerkiksi tuotekehityksessä. Suurten datamassojen analysoinnilla haetaan selkeitä etuja ja hyötyjä kuten uusia liiketoiminta-mahdollisuuksia, kustannussäästöjä, yksityiskohtaisempia ennusteita tulevasta toiminnasta ja tarkempia tietoja päätöksen teon tueksi.

Big data mahdollistaa lisäarvon tuottamisen asiakkaille ymmärtämällä asiakkaan tarpeita paremmin ja luomalla mielenkiintoisia, juuri heille sopivia personoituja ja kohdennettuja

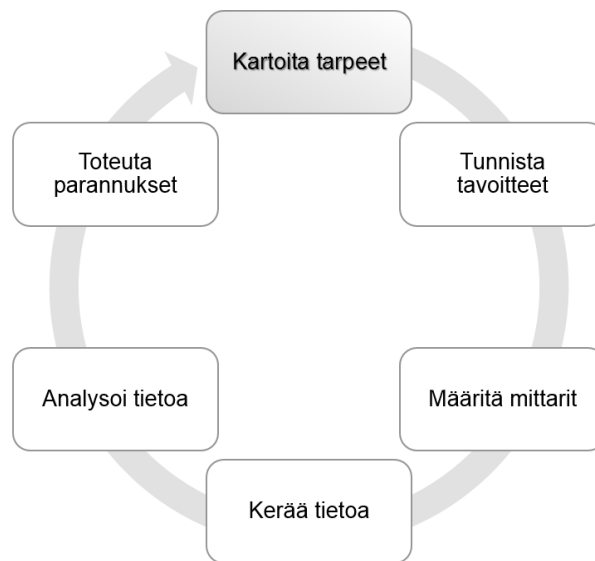
sisältöjä, tuotteita ja palveluita. Big data mahdollistaa paremman, henkilökohtaisen asiakaspalvelun ja -kokemuksen. (Liikenne- ja viestintäministeriö 2014, 8.) Data Science Series (2012) kiteyttää artikkelissaan ”Ten Practical Big Data Benefits” big datan oleelliset hyödyntämisen mahdollisuudet: kuluttajien profilointi, liiketoiminnan optimointi, yksityiskohtaisen tiedon saaminen päätöksenteon tueksi, uusien tuotteiden ja palveluiden löytäminen, parempi vuorovaikutus asiakkaiden kanssa ja ennustaminen.

3 Tiedosta päätöksiin

Englantilaisen Internet palveluja tuottavan Netcraft-yrityksen tutkimuksen mukaan verkkosivujen määrä maailmalla on lähes miljardi (Netcraft 2015). Sosiaalisen median kanavilla kuten Twitter on 255 miljoonaa, LinkedIn 187 miljoonaa ja Facebook 1.39 miljardia aktiivista käyttäjää. Twitteristä maailmalle lähtee joka päivä yli 500 miljoonaa twiittausta. (Digital Insights 2014.) Maailmassa luodaan päivittäin 2,200 petatavua tietoa. Tämä luku sisältää muun muassa 2 miljoonaa hakua Googlesta jokaisen minuutin aikana, 4000 tuntia videoita ladattuna YouTubeen jokaisen tunnin aikana sekä 144 miljardia sähköpostiviestiä lähetettynä ympäri maailmaa. (Garside ja Cox 2013, 4.) 34 % markkinoijista kerää tietoa asiakkaiden verkkokäyttäytymisestä ja 37 % ei kerää tietoa laisinkaan. Vain 22 % yhdistää asiakastietoa ja kanavatietoa. (Fonecta 2015.) Vaikka analytiikka ja tiedolla johtaminen ovat vielä lapsen kengissä Suomessa, mahdollisuudet kävijäseurantatietojen hyödyntämiselle on tiedostettu eri toimialoilla. Verkkoanalytiikalla haetaan yhä enemmän ja enemmän uusia näkökulmia tiedolla johtamiseen ja päätöksen teon tueksi.

Kansainvälisen Verkkoanalytiikka Yhdistyksen (2008, 3) virallinen määritelmä verkkoanalytiikalle on, että sen avulla kerätään ja mitataan tietoa, analysoidaan ja raportoidaan internetliikennettä verkon käytön ymmärtämiseksi ja tehostamiseksi. Sosiaalisen median myötä verkkoanalytiikka käsitteen rinnalle on tullut uusi käsite digitaalinen analytiikka, joka Kansainvälisen Digitaalinen Analytiikka Yhdistyksen (aikaisemmin Kansainvälinen Verkkoanalytiikka Yhdistys) määritelmän mukaan tarkoittaa suorituskyvyn parantamista ja tulevaisuuden ennustamista ymmärtämällä historiaa. Digitaalisella tässä kohtaa tarkoitetaan tietojen keräämistä hyödyntäen interaktiivisia kanavia kuten sosiaalista mediaa. (DAA 2015.) Liikenteen mittaaminen sivuilla, mielenkiinnon herättävien tuotteiden ja palveluiden poimiminen sivuilta, kampanjoiden onnistumisen mittaaminen, potentiaalisten asiakkaiden määrittäminen ja kävijöiden hakusanojen jäljittäminen ovat esimerkkejä verkkoanalytiikan käyttökohteista ja mitattavista asioista. Tuloksia analysoimalla yritykset saavat arvokasta tietoa päätöksen teon tueksi.

Verkkoanalytiikka jaetaan kuuteen erilliseen vaiheeseen (kuvio 4). Jokainen vaihe on oleellinen osa onnistunutta verkkoanalytiikkatoteutusta. Peräkkäiset vaiheet muodostavat yhdessä jatkuvan kehitysprosessin, jota toistetaan yhä uudestaan tulosten saavuttamiseksi ja parantamiseksi. (Kaushik 2010, 63.) Prosessin uusi kierros alkaa tarpeiden kartoittamisesta ja päättyy parannuksien, tässä tapauksessa uusien mittareiden jalkauttamiseen ja tarpeettomien mittareiden seurannan päättämiseen. Kaushik (2010, 64) suosittelee kirjassaan kehitysprosessin suorittamista säännöllisesti, vähintään neljä kertaa vuodessa.



Kuvio 4. Verkkoanalytiikan viitekehys

3.1 Tarpeiden kartoittaminen

Jokaisen yrityksen tarpeet verkkoanalytiikalle ovat erilaiset. On tärkeää, että tarpeet kartoitetaan ennen välineen valintaa ja mittareiden määrittelyä. Kaushikin (2010, 20–21) mukaan on oleellista, että tarpeiden kartoittamiseksi yritys miettii kolmea asiaa: tarvitaanko yrityksessä raportointia vai analysointia, onko yrityksen vahvuus tietotekniikassa vai liiketoiminnassa ja kolmanneksi onko tarkoitus selvittää käyttäjien klikkauspolkuja vai toteuttaa tiedon analysointia tarkemmalla tasolla. Raportointia käytetään yrityksen tiedon kokoamiseen, yhdistelyyn ja jakeluun. Raportointi kertoo mitä on tapahtunut. Analysoinnin avulla kokonaisuudet pilkotaan pienempiin osa-alueisiin, jotta asiasta saadaan parempi ymmärrys. Analytiikan avulla haetaan vastauksia siihen miksi asioita on tapahtunut.

Verkkoanalytiikkavälineet tuottavat valmiita raportteja keskiarvoista ja yksittäisistä mittareista, mutta monesti raportit sellaisenaan eivät palvele yrityksiä. Jotta valmiita raportteja

voidaan muokata, tulee yrityksen miettiä minkälaista ja mihin tarkoitukseen tietoa tarvitaan. Yrityksen on tarpeellista miettiä heti alkuun strategia analytiikkaa varten. Jos yritys haluaa jäljittää käyttäjien klikkauspolut, ovat vaatimukset verkkoanalytiikalle suppeammat, kuin jos tavoitteena on rakentaa johdettuja mittareita yhdistelemällä tietoa eri lähteistä. Jos yrityksen vahvuus on liiketoiminnassa, eikä varsinaista omaa IT-organisaatiota ole, on syytä harkita verkkoanalytiikkapalveluja tuottavan konsulttiyrityksen käyttöä. Oikean välineen valinta on oleellinen osa onnistunutta verkkoanalytiikkatoteutusta. Kaushikin (2010, 20) mukaan seitsemän kymmenestä yrityksestä epäonnistuu hankkeissa, jos asiantuntija ei ole mukana tuote-evaluoinnissa ja verkkoanalytiikan toteutuksessa.

3.2 Tavoitteiden tunnistaminen

Seuraava vaihe viitekehyksessä on tunnistaa ja asettaa yrityksen tavoitteet kattamaan asetetut tarpeet. Tavoitteiden avulla pyritään parantamaan esimerkiksi yrityksen verkkosivujen toimintaa siten, että hyöty saadaan niistä irti mahdollisimman hyvin. Oleellista on määritellä tavoiteltavat muutokset, niiden tuomat hyödyt ja asettaa muutoksille mitattavat tavoitteet sekä tavoitteille priorisointi. Tavoitteet ovat hyvin erilaisia riippuen yrityksestä. Sähköisen kaupankäynnin parantamisen tavoitteena on kasvattaa myyntiä ja vähentää markkinoinnin kuluja. Kun taas esimerkiksi sisältsivuston tavoitteena on kasvattaa sivuston lukijoiden määrää ja heidän viettämänsä aikaa sivustolla. (Clay 2015.) Konkreettisenä esimerkkinä Clay (2015) listaa muutaman mittavissa olevan tavoitteen: 100 kappaletta potentiaalisia myyntitapahtumia kuukaudessa, 50 kappaletta uutiskirjeen tilausta kuukaudessa ja 100 verkkosivurekisteröintiä kuukaudessa.

3.3 Mittareiden määrittäminen

Mittareiden määrittäminen on oleellinen osa verkkoanalytiikkaa. Jotta analytiikasta saatua tietoa voidaan hyödyntää liiketoiminnan tukena, tulee yrityksen tunnistaa ja määritellä omat tarpeensa ja tavoitteensa mittarien lähtökohdaksi. Tavoitteet asettavat mittareille vaatimukset. Kävijäseurannan lisäksi jokaisen yrityksen on mietittävä mittareita niille osaluueille, joita halutaan jatkossa analysoida, kehittää ja parantaa verkkoanalytiikan avulla. Keskeiset käsitteet kävijäseurannalle ovat sivu, käynti ja kävijä. Nämä käsitteet luovat pohjan useimmille mitattaville asioille. Verkkoanalytiikan kävijäseurannan perusmittareita (kuvio 5) ovat käynnit ja kävijät (engl. visits, visitors) sekä vierailuimmat sivut (engl. page views). Lisäksi yleisesti käytettäviä mittareita ovat kävijöiden tulotapa sivustolle (engl. traffic sources), hakukoneissa käytetyt avainsanat (engl. keywords) ja poistumisprosentti sivustolta (engl. bounce rate). (Kansainvälinen Verkkoanalytiikka Yhdistys 2008, 3-5.)

MITTARI	Visits	Page views	Traffic sources	Referring keywords	Bounce rate
KUVAUS	Sivustolle tulevien vierailijoiden (uusien ja palaavien) määrä	Vierailluimmat yksittäiset sivut	Kävijöiden tulotapa	Käytetyt avainsanat	Sivustolta poistumisprosentti
TUNNUSLUVUT	Vierailijoiden määrä sekä liikkumistapa sivustolla.	Vierailijoiden määrä yksittäisillä sivuilla sekä suosituimmat sivut	Vierailijoiden saapuminen sivustolle suoraan tai esim. hakukoneiden (Google, Yahoo, Bing) kautta.	Avainsanojen, joita vierailijat käyttävät päätyäkseen sivustolle, tunnistaminen.	Vierailijoiden määrä, jotka käyvät yhdellä sivulla ja poistuvat.

Kuvio 5. Verkkoanalytiikan perusmittarit

Kävijämäärät

Kävijämäärät mittaavat sivustolle tulevien yksittäisten vierailijoiden määrää sekä heidän liikkumistapojaan sivustolla. Kävijämääristä voidaan erotella uusien vierailijoiden määrät (engl. new visitors) ja niiden vierailijoiden määrät, jotka palaavat sivustolle toistuvasti (engl. return visitors). Kansainvälinen Verkkoanalytiikka Yhdistys (2008, 8–12.) Vierailijamittarien avulla voidaan mitata esimerkiksi markkinoinnin, kampanjoiden sekä sesongin vaikutuksia sivuston kävijämääriin. Tunnusluvut antavat arvokasta tietoa verkkosivujen kehittämistä silmälläpitäen. Sivuston heikko näkyvyys vaikuttaa uusien kävijöiden saapumiseen sivustolle, kun taas palaavien vierailijoiden määrä kertoo mielenkiinnosta yrityksen palveluita ja tuotteita kohtaan.

Sivut

Vierailluimpien sivujen kävijämäärät mittaavat sivuston yksittäisten sivujen suosiota sekä sivustolla ja sivuilla vietettyä aikaa (Kaushik 2010, 44). Mittarin avulla saadaan tietoa siitä mitkä yksittäiset sivut sivustolla herättävät mielenkiintoa kävijöissä ja saavat käyttäjät palamaan sivuille. Mittarin avulla markkinointi voi kohdistaa kampanjat oikeisiin asioihin, ajankohtiin ja oikeille käyttäjille.

Tulotapa

Kävijöiden tulotapa sivustolle kertoo mitä lähdettä kävijät käyttävät saapuakseen sivustolle. Saapuvatko kävijät suoraan sivustolle, käyttämällä hakukoneita kuten Googlea, Yahooa tai esimerkiksi Bingiä vai verkkomainonnan kautta. Hakukoneissa käytettävät avainsanat lisäävät mittarin merkitystä. Sivustolle tulotavan lisäksi saadaan tietoa siitä mitä asioita käyttäjät etsivät päätyäkseen sivustolle. Mittarin avulla voidaan niin ikään mitata mainoskampanjoiden onnistumista ja niiden näkymistä kävijämäärissä. Verkkomainonta ohjaa

liikenteen sivustolle ja kampanjoiden aikana sivuston kävijämäärän tulisi luonnollisesti kasvaa. (Kaushik 2010, 86.)

Poistumisprosentti

Välitön poistumisprosentti mittaa kävijöiden määrän yhdellä sivulla ennen heidän poistumista sivustolta. Syitä poistumiseen voi olla useita kuten sivustojen ulkoasu, sivuston käytettävyys tai tietty ajankohta. Syiden löytyminen on haasteellista ja vaatii myös muiden kävijäseurannan mittareiden analysointia sekä niiden yhdistämistä poistumisprosenttiin. Poistumisprosenttiin voi vaikuttaa muun muassa sivustolle tulotapa. (Kaushik 2010, 51–53.)

3.4 Tiedon keruu

Verkkoanalytiikkavälineet tukevat kahta erilaista tiedonkeruumenetelmää. Menetelmiä kutsutaan lokitiedosto- ja JavaScript -menetelmiksi. Kumpikin menetelmä on yleisesti käytössä ja kummastakin menetelmästä on löydettävissä sekä hyviä että huonoja puolia (taulukko 2). Lokitiedostojen puolesta puhuu niiden edullisuus kun taas Java-skriptit mahdollistavat reaaliaikaisen seurannan. Huonojen ominaisuuksien poistamiseksi ja luotettavan tiedon varmistamiseksi kävijäseurantaohjelmistoihin on myös yhdistelty näitä kahta tiedonkeruumenetelmää.

Taulukko 2. Tiedonkeruumenetelmien vertailu

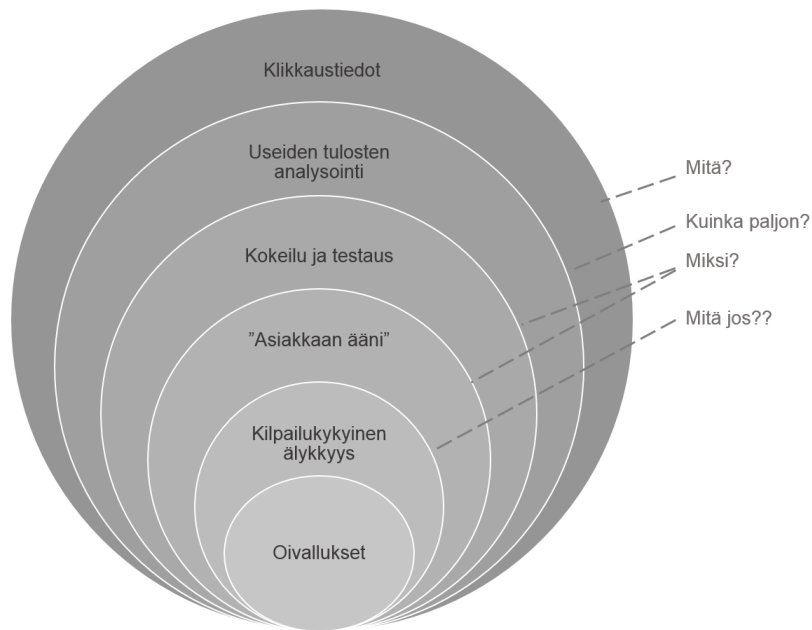
LOKITIEDOSTOT	JAVA-SKRIPTIT
<ul style="list-style-type: none">• Vähäinen muutoksien tarve• Kertamaksu, mahdolliset lisäkustannukset• Halvempi• Mahdollistaa hakukoneen indeksoinnin	<ul style="list-style-type: none">• Vaatii koodin jokaiselle sivulle• Kuukausiveloitus• Käytetympi• Reaaliaikainen

Palvelinlokityiedostot ovat perinteinen tapa kerätä tietoa käyttäjien toiminnoista yrityksen sivustolla. Palvelin tallentaa lokitietoa kaikista palvelimille tehdyistä pyynnöistä. Lokityiedostoihin kerätään huomattava määrä tietoa, jota voidaan hyödyntää esimerkiksi kävijäseurannassa. Lokit sisältävät tietoa tapahtumista ja niiden ajankohdista. Lokityiedostojen hyödyntäminen sellaisenaan on haasteellista, koska tietoa on paljon ja eikä se ole yksinkertaisessa luettavassa muodossa. Lokitytietojen lukeminen tapahtuu parhaiten verkkoana-

lytiikkaohjelmien avulla, jotka osaavat tulkita lokitietoa ja muuttaa tiedot tunnusluvuiksi. (Clay 2015.) Toinen tapa tiedon keräämiselle on Java merkin "Page tag" lisääminen jokaiselle seurattavalle verkkosivulle. Merkki sisältää tietoa, kuten mikä sivu on kyseessä ja onko kävijä käynyt sivustolla aikaisemmin. Kävijän saapuessa verkkosivulle JavaScript havaitsee käynnin ja siirtää tiedot verkkoanalytiikkapalvelimelle. (Clay 2015.)

3.5 Tiedon analysointi

Kun pohjatyö on tehty hyvin: ongelmakohdat on tunnistettu, päämäärät on asetettu, väline on valittu ja tietoa on saatavilla, voidaan verkkoanalytiikkaa alkaa hyödyntämään. Verkkoanalytiikkavälineillä tietoa voidaan analysoida monella tasolla. Siksi oikean välineen hankinta on oleellista, vaihtoehtoja tulee miettiä jo tavoitteita asetettaessa. Kaushik kuvaa kirjassaan analysoinnin tasot (kuvio 6), listaa mahdolliset tavat näiden tasojen saavuttamiseksi sekä antaa vaihtoehtoja oikean välineen valintaan yrityksen koosta riippuen (taulukko 3). Mitä syvempi taso sitä yksityiskohtaisempiin kysymyksiin saadaan vastauksia: kuinka paljon, miksi ja mitä jos. Nämä avainkysymykset Kaushik kohdistaa analysoinnin tasoihin: klikkaustiedot, useiden tulosten analysointi, kokeilu ja testaus, "asiakkaan ääni", "kilpailukykyinen älykkyys" ja oivallukset. Klikkaustieto on verkkoanalytiikan perustietoa, jonka avulla saadaan vastauksia kysymykseen mitä. Sen avulla voidaan mitata kävijöiden, käyntien ja näyttökertojen määrää, mutta analysointia voidaan syventää poimimalla kävijäseurannasta tietoa kuten suosituimmat sivut, kävijöiden tuloreitit, välitön poistumisprosentti ja sivuilla vietetty aika. Kaushik suosittelee analysoinnin tasoja käytettäväksi sekä pienissä, keskisuurissa että suurissa yrityksissä. Pienissä yrityksissä riittää tasot 1-3, suuremmissa yrityksissä suositellaan kaikkien tasojen hyödyntämistä. (Kaushik 2010, 7-10.)



Kuvio 6. Verkkoanalytiikan analysoinnin tasot ja avainkysymykset eri tasoille (Kaushik 2010, 7)

Taulukko 3. Suositukset analysoinnin tasoista erikokoisissa yrityksissä (Kaushik 2010, 13)

Analysoinnin tasot	Pienet yritykset	Keskisuuret yritykset	Suuret yritykset
KLIKKAUSTIEDOT	1	2	3
USEIDEN TUOTTEIDEN ANALYSOINTI	2	1	2
KOKEILU JA TESTAUS		4	4
"ASIAKKAAN ÄÄNI"	3	3	1
KILPAILUKYKYINEN ÄLYKKYYS			5
OIVALLUKSET			6

3.6 Parannuksien toteuttaminen

Jatkuva ja säännöllinen sivuston tuloksien seuranta ja asetettujen tavoitteiden ja mittareiden hienosäätö ovat lähtökohdat parannuksien toteuttamiselle. Mittarien raportoinnin lisäksi tarvitaan jatkuvaa analysointia. Yritykset, jotka läpikäyvät olemassa olevaa tietämystä ja tekevät muutoksia strategioihin, tavoitteisiin, tiedon keruuseen ja mittauksiin ovat lähempänä asiakaskeskeisyyttä kuin muut yritykset. Peruslähtökohta parannuksien toteut-

tamiselle on saada parempi ja nopeampi ymmärrys asiakkaiden käyttäytymisestä kuin kilpailijat ja hyödyntää sitä omassa liiketoiminnassaan.

Tohtori Andrew Jennings listaa artikkelissaan (FICO 2015) ”When Is Big Data the Way to Customer Centricity?” neljä ohjetta, joiden avulla liiketoiminta saa big datasta ja analytiikasta parhaan hyödyn irti. Suunnittele ja automatisoi kokeiluja, joiden avulla mahdollistat ennustamisen. Analysoi ja opi asiakkaiden käyttäytymistä ”lennosta”. Tasapainota tekniikka, inhimillinen tietämys ja kokemus sekä käännä jokainen asiakaskontakti mahdollisuudeksi.

4 Case: Big datan ja analytiikan hyödyntäminen

Verkkoanalytiikan tuottama sosiaalisesta mediasta kerätty tieto ei yksinään ole riittävä yrityksille tukemaan päätöksentekoa. Se täydentää rakenteellista tietoa erinomaisesti, ja yhdistämällä se laaja-alaisesti ja monipuolisesti rakenteelliseen tietoon, voidaan sitä hyödyntää liiketoiminnan johtamisen tukena. Rakenteellisen tiedon yhdistämisellä rakenteetomaan tietoon mahdollistetaan kattavampi pohja tiedolla johtamiseen. ”Hyöty ei aina ilmene heti ja ole saatavissa suoraviivaisesti, mutta tulevaisuudessa omaa dataa esimerkiksi muiden reaaliaikaisten tietolähteiden kanssa yhdistelemällä saatetaan saavuttaa lisäarvoa” (Liikenne- ja viestintäministeriö 2014, 14).

Markkinoilla on useita kymmeniä big data ja analytiikkaohjelmia, sekä ilmaisia avoimen lähdekoodin ratkaisuja että maksullisia ratkaisuja. Useat suomalaiset yritykset tekevät tällä hetkellä mittavia kehityshankkeita big datan, ennakoivan analytiikan sekä yrityksen oman tietopääoman hyödyntämiseksi tiedolla johtamisen tueksi. Invenco on mukana Digile Oy:n ja Innovaatorahoituskeskus Tekesin Need For Speed -yhteistyökonsernissa ohjelmassa. Ohjelman tavoitteena on luoda yrityksille perusta, joka mahdollistaa menestymisen digitaalisessa taloudessa. Ohjelma on nelivuotinen, 2014–2017 ja sitä toteuttamassa ovat suomalaiset erikokoiset ohjelmistoyritykset. (Digile 2015.)

Invenco suunnitteli ja toteutti ohjelman yhteydessä valikoimaan analytiikkatyökalun. Sen avulla yrityksissä hajallaan oleva rakenteellinen tieto ja ulkopuolelta saatua verkkoanalytiikkatietoa voidaan kerätä, yhdistää, raportoida ja esittää erilaisina asiakkaan määritteleminä mittareina. Analytiikkatyökalu on toistaiseksi pilottiversio, joka sisältää rajallisen määrän toimintoja. Tavoitteena Invencolla on tuotteistaa konfiguroitava työkalu, jonka avulla saavutetaan lisäarvoa asiakkaille yhdistämällä eri tietolähteitä ja tuottamalla juuri oikeanlaisia mittareita tukemaan tiedolla johtamista asiakasorganisaatiossa.

Analytiikkatyökalu toteutettiin pilotoiden yhteistyössä Invencon liiketoimintatiedon hallintaan erikoistuneiden konsulttien ja kotimaassa elintarviketoimialalla toimivan referenssi-asiakkaan kanssa vuoden 2014 loppupuoliskolla ja vuoden 2015 alkupuoliskolla. Pilotin toteutuksessa huomioitiin referenssiasiakkaan liiketoiminta sekä verkkoanalytiikkaviitekehys. Aikaisemmin opinnäytetyön teoriaosuudessa kuvattu verkkoanalytiikan viitekehys tuki analytiikkatyökalun ideointia, suunnittelua sekä toteutuksen pilotointia.

4.1 Analytiikkatyökalulle asetut tarpeet ja tavoitteet asiakasnäkökulmasta

Työ aloitettiin haastattelemalla referenssiasiakasta. Haasteena asiakkaalle oli ymmärtää kuinka tämä erilaisissa muodoissa oleva tieto tulee kerätä, tallentaa ja liittää yrityksen tietovarastoissa olevaan tietomassaan. Oleellista tavoitteiden kannalta on, että yrityksessä ymmärretään tiedon laajempi käyttö yhdistäen sekä sisäisiä järjestelmiä että ulkoisia lähteitä. Jalostetun tiedon hyödyntäminen vaatii yritykseltä tilastollista osaamista, ymmärrystä mihin tietoa voidaan käyttää eli mihin haetaan vastauksia. Haasteena nähtiin niin ikään teknisen osaamisen puuttuminen sekä puutteellinen integraatio olemassa olevan raak tiedon ja yrityksen sisäisten järjestelmien välillä.

Keskusteluissa esille nousi tarve saada kokonaiskuva yrityksen tiedoista yhdistämällä verkkoanalytiikka tietoa ja tietovarastoissa olevaa rakenteellista tietoa yhdeksi tiedolla johtamisen sovellukseksi. Toisena tavoitteena oli mahdollistaa yritykseen liittyvien sosiaalisen median trendien ja keskustelujen seuraaminen reaaliaikaisesti. Esimerkkinä miten uuden tuotekampanjan julkistaminen näkyy keskusteluina sosiaalisessa mediassa. Kolmantena tavoitteena pilotoinnissa oli tuottaa referenssiasiakkaalle perinteisen raportoinnin rinnalle tietojen visualisointia sekä kerätystä ja tallennetusta että reaaliaikaisesta tiedosta. Kaikkien edellä mainittujen tavoitteiden lisäksi analytiikkatyökalun tulisi olla helppokäyttöinen ja yksinkertainen, muunneltava ja joustava, nopea ja ajantasainen eikä sen käyttö saisi vaatia käyttäjiltään teknistä syväosaamista.

Lopputuloksena syntyivät verkkoanalytiikan tarjoaminen perusmittareiden lisäksi vaatimukset yksilöllisistä referenssiasiakkaan liiketoiminnalle koostettavista mittareista. Liiketoiminnan näkökulmasta tiedon analysointitarpeiksi listattiin muun muassa valitun tuotteen myyntimäärien vastaavuus sosiaalisessa mediassa saatuihin trendeihin, liikevaihdon kehitys ja sen muutoksien analysointi sosiaalisen median trendien mukaisesti, markkinointi kampanjoiden vaikutus sosiaalisessa mediassa esiintyviin trendeihin, kilpailijoiden trendien analysointi, kysynnän ennustaminen perustuen trendeihin ja analysointi kuinka hyvin kysynnän ennustaminen on toteutunut. Trendillä tässä kohtaa tarkoitetaan valitun tuotteen esiintymistä verkossa. Lisäksi tarpeina nähtiin kerätä tietoa, johon yrityksen tulisi reagoida helposti ja nopeasti kuten muun muassa positiiviset ja negatiiviset mielikuvat tuotteista ja niihin liittyvät keskustelut, sään vaikutus myyntiin sekä kilpailijan tuotteista tietojen kerääminen.

4.2 Liiketoimintalähtöiset mittarit perustuvat asetettuihin tavoitteisiin

Tavoitteiden pohjalta suunniteltiin mittarit, jotka pyrittiin pitämään mahdollisimman yleisluonteisina. Tavoitteena oli, että analytiikkatyökalun pilottiversion mittarit soveltuisivat jat-

kossa myös muiden toimialojen käyttöön sekä herättäisivät keskustelua liiketoimintakoh-
taisia mittareita suunniteltaessa. Kuten lähteet, myös mittarit luokiteltiin neljään kokonai-
suuteen. Ne perustuvat yrityksen omiin nettisivuihin, sosiaaliseen mediaan, kaikkiin julki-
siin sivuihin sekä edellä mainittujen tietojen yhdistämiseen yrityksen omaan rakenteellisen
tiedon esimerkiksi tietovarastotiedon kanssa.

Kävijäseurannan mittarit

Verkkosivut toimitetaan lähes aina sisältäen kävijäseurannan. Luvut antavat yritysjohdolle
yritystä koskevaa, vertailukelpoista ja ajantasaista tietoa. Kävijäseurannan merkitys on
merkittävä sivuston parantamisen kannalta, mutta analysoituna ne antavat arvokasta tie-
toa päätöksen teon tueksi. Analytiikkatyökaluun referenssiasiakkaalle valittiin oleellisim-
mat tunnusluvut: kävijämäärien seuranta sekä uusien että palaavien asiakkaiden osalta,
sivustolle tulotapa, eteneminen sivuilla, suositut sivut sekä sivuilla tehdyt haut. Lisäksi
sivustolta haluttiin seurata kävijän maantieteellistä sijaintia sekä kerätä tietoa kävijöiden
käyttämistä selaimista ja fyysisistä laitteista.

Sosiaalisen median mittarit

Sosiaalinen media tuottaa liikennettä yrityksen www-sivuille. Sen merkitys nykypäivänä on
suuri ja siksi kävijäseurannan mittareita haluttiin rakentaa myös tarkastelemaan eri sosi-
aalisen median kanavien tuottamaan liikennettä. Referenssiasiakkaalle luotiin mittarit,
jotka raportoivat niiden kävijöiden määrän, jotka päätyvät sivustolle sosiaalisen median
kautta ja minkä sosiaalisen median kanavan kautta. Käynnin syy kertoo mikä sai kävijän
tulemaan sivuille esimerkiksi omat julkaisut, mainokset tai bannerit. Sivustolla mitataan
myös sosiaalisessa mediassa tapahtuvien julkaisujen, mainoskampanjoiden tai sesonki-
jaksojen vaikutusta sivustolla tapahtuvaan liikenteeseen. Lisäksi raportoidaan mitä ja
kuinka paljon sosiaalisessa mediassa keskustellaan kilpailijoista verrattuna referenssiyri-
tykseen.

Julkisiin www-sivustoihin perustuvat mittarit

Hakukoneet kuten Google, Yahoo ja Ask ovat yleisimpiä tapoja päätyä halutuille sivuille.
Referenssiasiakkaan analytiikkatyökaluun toteutettiin mittarit, joiden avulla mitataan julki-
silla www-sivuilla käyttäjien kiinnostuksen kohteita. Mitataan hakusanojen kuten tietyn
tuotteen tai palvelun esiintymistä blogeissa ja keskustelupalstoilla. Lisäksi haetaan tietoa
siitä mitä ja kuinka paljon mediassa esiintyy arvosteluita tai arvioita yrityksestä itsestään
tai kilpailijoista, ja minkälaista mielikuvaa: negatiivista, positiivista tai neutraalia ne herät-
tävät julkisissa keskusteluissa, artikkeleissa ja blogeissa.

Koostetun tiedon mittarit

Oleellisena lisänä referenssiasiakkaalle määriteltiin mittareita, jotka perustuvat tietojen yhdistämiseen. Yritysten tietojärjestelmissä on huomattava määrä rakenteellista tietoa, joita yhdistämällä ja vertailemalla eri lähteistä saatuun tietoon, saadaan paljon arvokasta tietoa yrityksen tilasta, kokonaisuudesta ja kehityskohteista. Koostetut mittarit ovat yleisesti yrityskohtaisia ja vaativat kohdetiedon tuntemusta sekä suunnittelua. Referenssiasiakkaalle määriteltiin mittareita, joiden avulla muun muassa sään, maantieteellisten erojen, verkkosivujen kävijämäärän, markkinoinnin, sosiaalisen median aktiivisuuden sekä mielikuvan vaikutusta myyntiin voitiin seurata. Lisäksi haluttiin mitata trendien vaikutusta ja osuvuutta tuotteiden myyntiin ja liikevaihtoon. Myös kilpailevien tuotteiden trendejä ja niiden vaikutusta myyntiin katsottiin tarpeelliseksi seurata.

4.3 Analysoitava tieto kerätään valituista lähteistä

Tiedonkeruu mittareille tapahtuu käyttäjän valitsemista lähteistä, jotka ovat täysin määriteltävissä analytiikkatyökalussa. Lähteinä ovat yrityksen omat www-sivut, sosiaalisen median kanavat, kaikki julkiset sivut esimerkkeinä tilastokeskus, ilmatieteenlaitos erilaiset blogit ja keskustelupalstat sekä yrityksen oma rakenteellinen tieto kuten asiakas-, tuote-, palvelu- ja hintatiedot tai esimerkiksi liikevaihto tuotteittain tai alueittain tai ennusteet tuotteittain tai alueittain.

4.4 Mittarit visualisoidaan reaaliaikaisina tai staattisina raportteina

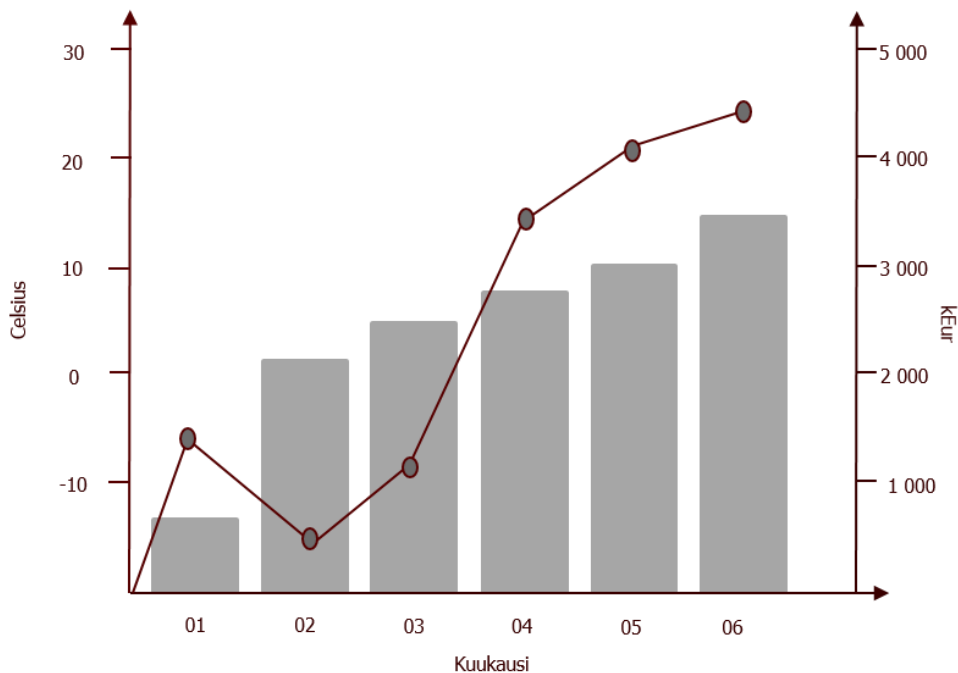
Kuten jo aikaisemmin todettiin, ei tietojen kerääminen ja tallentaminen itsessään anna yritykselle ymmärrystä tiedon sisällöstä. Tietoa pitää jalostaa ja muokata esitettävään muotoon. Analytiikkatyökaluun mittarien muodostus kerätystä datasta toteutettiin sekä määrällisenä että laadullisena tutkimuksena. Määrällisessä tutkimuksessa käytetään täsmällisiä ja laskennallisia menetelmiä ja pääasiassa ne muodostetaan laskemalla lähdetietoja (Jyväskylän yliopisto 2015a).

Referenssiasiakkaan vaatimuksena oli raportti, jossa näkyvät yrityksen sivustolle saapuneet kävijät tulopaikan ja selaimen mukaisesti (kuvio 7). Tiedot raportille haetaan sosiaalisesta median kanavaista ja yrityksen omilta verkkosivuilta verkkoanalytiikka hyödyntäen.

	Kävijämäärä		Tullut sivustolle						Selain					Keskimääräinen aika viivytty sivulla (min.)
	Vierailut	Vierailijat	Google	Twitter	Facebook	LinkedIn	Etusivulta	Muu/ei tiedossa	IE	Firefox	Chrome	Safari	Muu/ei tiedossa	
Etusivu	1234	1111	567	230	5	23	320	89	395	321	259	234	25	00:08:32
Sivu 1	345	300	45	1	2	5	290	2	110	90	72	66	7	00:01:10
Sivu 2	123	100	3	0	0	0	120	0	39	32	26	23	2	00:02:03
Sivu 3	4321	3456	1221	2345	398	14	340	3	1383	1123	907	821	86	00:05:15
Sivu 4	10	10	0	0	0	0	10	0	3	3	2	2	0	00:00:26
Sivu 5	789	650	432	98	29	0	230	0	252	205	166	150	16	00:01:57

Kuvio 7. Kävijämäärä tulopaikan ja selaimen mukaan, tietolähteinä raportille omat www-sivut ja sosiaalinen media

Vaatimuksena oli myös raportti, jossa kuluvan vuoden myyntiä verrattiin pääkaupunkiseudun keskilämpötiloihin (kuvio 8). Tässä raportissa tietolähteenä käytetään ilmatieteen laitoksen julkista sivustoa sekä yrityksen myyntitietoja.



Kuvio 8. Kuluvan vuoden myynti vs. kuukauden keskilämpötila pääkaupunkiseudulla, tietolähteenä julkiset verkkosivut ja yrityksen rakenteellinen data

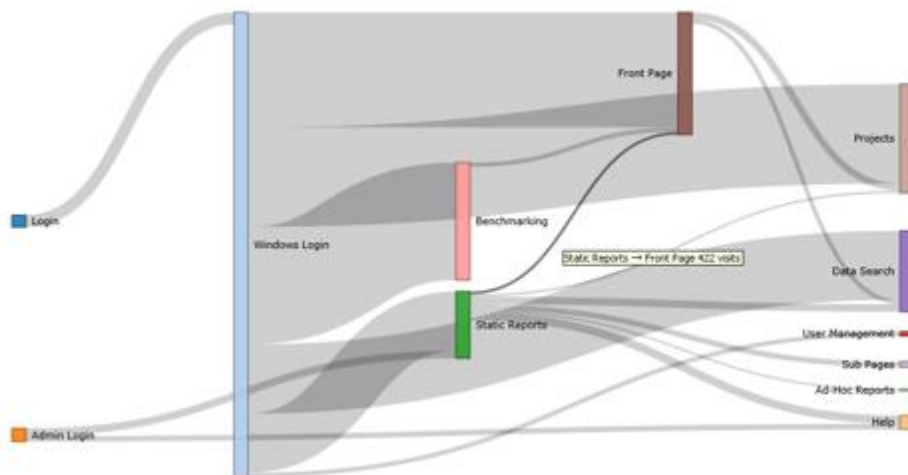
Tuotteiden suhteutettu lukumäärällinen esiintyminen esitettiin staattisena raporttina (kuvio 9), jossa lähteenä olivat sekä yrityksen rakenteellinen tieto että julkiset verkkosivut, avoin tieto.



Kuvio 9. Tuotteiden suhteutettu lukumäärällinen esiintyminen, tietolähteenä julkiset verkkosivut.

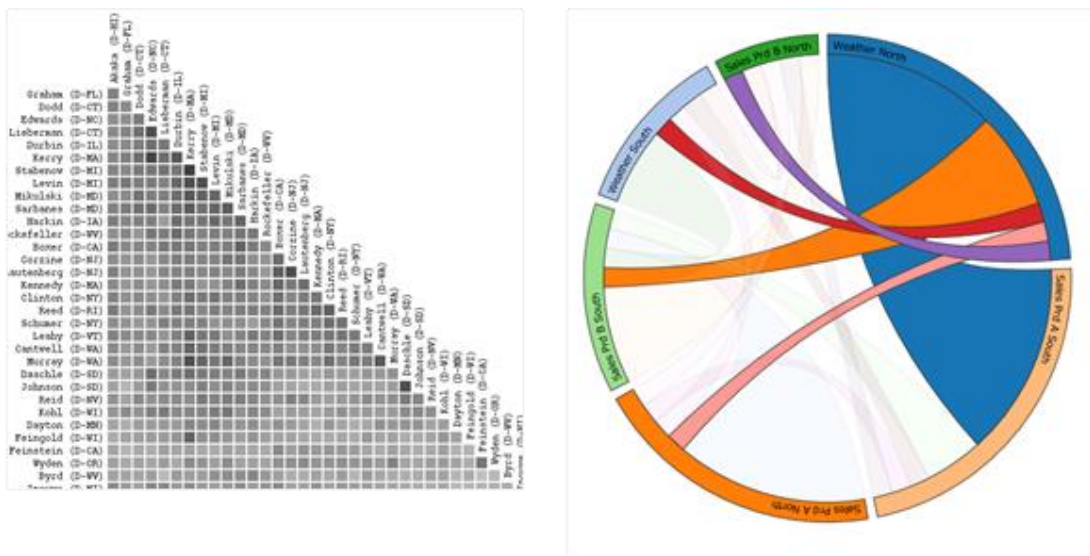
Laadullisessa tutkimuksessa pyritään ymmärtämään mistä ilmiössä on kyse. Laadulliseen tutkimukseen tarvitaan ilmiön merkityksen tai tarkoituksen tarkempaa selvittämistä. (Jyväskylän yliopisto 2015b.) Laadullinen tutkimus vaatii sisällön tunnistamista, kuten esimerkiksi tweet-viestin tunnistamista positiiviseksi tai negatiiviseksi. Tällainen tieto voidaan koneellisesti muodostaa käyttämällä tekstin arvoanalyysiä eli oppivaa logiikkaa, jota hyödyntämällä voidaan tunnistaa tietyntyyppiset viestit. Referenssiasiakkaalle kerättiin sosiaalisen median kanavista sekä positiiviset että negatiiviset keskustelut ja mielikuvat yrityksestä itsestään ja valituista tuotteista.

Analytiikkatyökaluun toteutettiin esimerkkinä mittarit myös verkkosivujen lokien seulonnalle sekä eri tietolähteiden yhteneväsyydelle. Verkkosivujen lokien seulonnalla löydetään yksittäiset polut ja summataan ne eri ryhmittelyihin perustuen esimerkiksi maahan, kaupunkiin tai ajanjaksoon. Lokien seulonnan avulla voidaan etsiä kiinnostuksen kohteet ja sivut, jotka eivät esimerkiksi saa huomioita. Sen avulla voidaan parantaa sivuston käytettävyyttä. Esimerkissä tulos esitetään Sankey-kuvaajan avulla (kuvio 10), jolla yleisesti esitetään virtauksia.



Kuvio 10. Sankey-kuvaaja navigointivirrasta, verkkopalveluiden lokien seulonta

Eri tietolähteiden välistä yhteyttä mitataan laskemalla kahden tai useamman tietolähteen välinen korrelaatio. Mittarin avulla voidaan havaita syy-seuraussuhteita tietolähteiden välillä ja asioita, joilla ei aiemmin tiedetty olevan liitoksia toisiinsa. Esimerkissä tulos esitetään matriisina tai Chord-kuvaajana (kuvio 11).



Kuvio 11. Korrelaatiomatriisi ja Chord-kuvaaja eri tietolähteiden välisestä yhteneväisyyksistä

Mittarien visualisointia tehdään sekä reaaliaikaisena että staattisina raporteina. Visualisoinnilla haettiin paremman ja yksityiskohtaisemman kuvan muodostamista mittareista, erilaisten trendien esittämisestä, raportoivan tiedon vertailua sekä erilaisten syy-seuraussuhteiden esille tuomista. Referenssiasiakkaan staattiset raportit toteutettiin nykyisillä Business Intelligence(BI) -raportointivälineillä. Summatasoisien verkkoanalytiikka-

datan siirtäminen yrityksen nykyiseen tietovarastoon mahdollistaa tiedon raportoimisen entuudestaan tutuilla välineillä. Näin kynnys raportoimiseen ja uuden tiedon hyödyntämiseen katsottiin olevan mahdollisimman matala.

4.5 Analytiikkatyökalun toiminnalliset kokonaisuudet

Analytiikkatyökalu jakautuu toiminnallisiin kokonaisuuksiin: tiedon keruu, mittarien tuottaminen, mittarien raportointi ja esitys (kuvio 12). Käyttäjän näkökulmasta analytiikkatyökalu on käyttöliittymä big data-järjestelmälle, jossa käyttäjä valitsee tiedonkeruun tietolähteet halutuista sosiaalisen median kanavista sekä mahdolliset yrityksen sisäiset järjestelmät, johon kerättyä tietoa halutaan yhdistää. Lisäksi käyttäjä valitsee tarkasteltavat mittarit sekä mittareiden tallennuspaikan jälkikäsitteilyä ja arkistointi varten. Tiedonkeruun ja mittareiden muodostumisen jälkeen käyttäjä voi siirtyä analysointipalveluun, jossa kerättyä tietoa voidaan analysoida summatasolla tai yksityiskohtaisesti. Käyttäjä voi seurata tilannetta yksityiskohtaisesti reaaliajassa analytiikkatyökalun valvonta- ja hallintaosiossa.



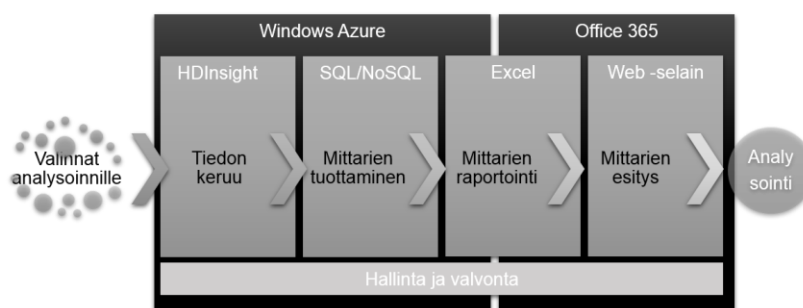
Kuvio 12. Analytiikkatyökalun toiminnalliset kokonaisuudet

4.6 Valittu tekniikka tukee analytiikkatyökalun toiminnallisuutta

Analytiikkatyökalu toteutettiin Microsoftin Windows Azurella pilviympäristössä hyödyntäen sen valmiita palveluja, erityisesti HDInsight-palvelua, joka on kehitetty big data tarpeisiin. HDInsight on Hadoop-ratkaisu, jonka avulla suuria määriä strukturoimatonta tietoa voidaan käsitellä ja analysoida tehokkaasti. Hadoop-ohjelmistokehys valittiin, koska se tarjoaa edullisen ja toimintavarman tiedon tallennuksen, tehokkaan rinnakkaisen prosessoinnin, skaalautuvalle yhdeltä serveriltä useille jopa tuhansille koneille sekä monipuolisen ekosysteemin lisäosien ja laajennuksien ansioista. Myös sen nopea käyttöönotto sekä alustariippumattomuus katsottiin eduksi. Toteutuksessa komponentit pyrittiin luomaan mahdollisimman yleiskäyttöisiksi, jotta toiminnot on tulevaisuudessa mahdollista ottaa käyttöön myös toisissa ympäristöissä. (Microsoft Azure 2015; Microsoft 2015.)

Analytiikkatyökalu sisältää tekniset komponentit prosessien hallintaan, tiedon keruuseen, mittarien raportointiin sekä niiden esittämiseen (kuvio 13). Prosessien hallintaa varten varattiin työasema, josta prosesseja voidaan ohjata, hallita ja valvoa. Hallinta sekä valvonta toteutettiin powershell-skripteillä sekä Windows Azure portaalin tarjoaminen mittarien avulla. Tiedon keruu toteutettiin hyödyntäen Hadoopin ydintä MapReduce-laskentamallia sekä Hive-tietovarastoinfrastruktuuria. MapReduce on tarkoitettu suurien tietomassojen rinnakkaiseen käsittelyyn ja analysointiin. Nimensä mukaisesti MapReduce jakautuu kahteen perustoiminnallisuuteen: map eli karttatoiminto ja reduce eli vähennystoiminto. MapReduce-mallin avulla iso työ voidaan pilkkoa pienempiin osiin ja suorittaa rinnakkain Hadoop-klusterin jokaisella jäsenellä. Map tarkoittaa datan pilkkomista ja palasen käsitteilyä, Reduce palasten summaamista. Kun nämä suoritetaan peräkkäin suoritusputkessa, saadaan massiiviset työt suoritettua tehokkaasti. MapReduce-laskentamallia hyödyntävät muun muassa Yahoo ja Facebook. Hadoopin päälle rakennettua Hiveä käytetään tiedon analysointiin sekä ad hoc -kyselyihin. Hiven avulla voidaan big dataa käsitellä SQL-tyyppisellä, HiveQL-kielellä. Hive-kyselyt generoituvat map- ja reduce -tehtäviksi. (Microsoft Azure 2015; Microsoft 2015.)

Mittareiden tuottaminen analytiikkatyökaluun toteutettiin SQL-HQL sekä PowerShell-skripteillä ja niiden raportointi ja esitys Office 365 -työkaluilla sekä Web-sivuilla. Kerätystä tietomassasta tehdään määrällistä sekä laadullista tutkimusta. Määrälliset mittarit kuten esimerkiksi suositut sivut, kävijän maantieteellinen sijainti muodostetaan pääasiassa laskemalla lähdetietoja. Laadullinen tutkimus vaatii sisällön tarkempaa tunnistamista, kuten esimerkiksi onko twiittaus positiivista vai negatiivista. Analytiikkatyökalussa tällainen tieto tunnistetaan käyttämällä oppivaa logiikkaa, joka toteutettiin hyödyntämällä ennakoivan analysoinnin työkalua Machine Learning. Koneoppiminen käyttää hyödykseen olemassa olevaa tietoa ennustaakseen tulevaa käyttäytymistä, tuloksia ja suuntauksia. (Microsoft Azure 2015; Microsoft 2015.)



Kuvio 13. Analytiikkatyökalun arkkitehtuuri ja käytetty tekniikka

5 Pohdinta

Opinnäytetyön tavoitteena oli selvittää, miten big dataa ja analytiikkaa voidaan hyödyntää tiedolla johtamisen tukena. Työn tietoperusta sisälsi kuvauksen siitä mitä tieto käsitteenä tarkoittaa. Missä ja mistä tietoa syntyy sekä miten tieto jalostuu päätöksenteon ja tiedolla johtamisen tueksi. Tietoperustan rajaaminen oli haastavaa, koska opinnäytetyön aihe oli laaja ja siihen liittyi useita näkökulmia. Mitä enemmän asiaan perehtyi, sitä enemmän ulottuvuuksia asian tiimoilta ilmeni. Tästä syystä myös aiheen ympärillä olevia käsitteitä oli paljon. Niiden kuvaaminen oli mielestäni oleellista, jotta opinnäytetyön empiirisen osuuden case –esimerkki olisi ymmärrettävissä. Yksi haasteellisimmista kuvattavista käsitteistä oli big data. Se on merkitykseltään hyvin monimuotoinen. Käsitteenä ja asiana sen voi ymmärtää monella tapaa. Sen merkitys voi muuttua hyvinkin erilaisiksi riippuen siitä minkälaisia asioita toimija pitää oleellisina. Tiedon määrä, monimuotoisuus, nopeus, tekniikka ja teknologiat luovat omanlaisensa haasteet ja mahdollisuudet käsitteen ympärillä. Kokonaisuudessaan mielestäni opinnäytetyön tietoperusta on kattava, ja se luo perustan opinnäytetyön empiiriselle osuudelle.

Analytiikkatyökalun tehtiin pilotointiprojektina Invencossa. Suunnittelu- ja toteutusvaihetta ohjasivat referenssiasiakkaalta saadut tavoitteet ja vaatimukset. Vaikka tavoitteet olivat hyvin konkreettisia, olisi asiakkaan jatkuva mukanaolo ollut suunnittelu- ja kehitystyön kannalta hyödyllistä. Lopputuloksen eli pilotin esittely asiakkaalle tuotti erinomaista palautetta, josta saimme hyviä ideoita jatkokehitystä silmällä pitäen. Analytiikkatyökalun pilottiversion kehitystyötä jatketaan Invencossa. Tavoitteena on työkalun tuotteistaminen. Analytiikkatyökalu ei tule tuotteena olemaan heti käyttöön otettava sovellus. Onkin oleellista miettiä toimintamalli miten analytiikkatyökalu on helposti ja nopeasti käyttöön otettavissa erilaisissa liiketoimintaympäristöissä. Etenkin johdettujen mittareiden osalta tavoitteet ja tarpeet voivat olla hyvinkin erilaisia riippuen asiakkaasta. Jatkokehitystyön rinnalla pilottia hyödynnetään Invencon myynnissä. Pilotti on tehokas myynnin väline, jonka avulla tavoitellaan uusia asiakkuuksia analytiikan alueelle. Jatkoa ajatellen pidän oleellisena erilaisiin tekniikkoihin ja teknologioihin tutustumisen ja niiden tuomien mahdollisuuksien hyödyntämisen analytiikkatyökalun jatkokehityksessä.

Empiirisessä osassa kuvattiin opinnäytetyön toimeksiantajan toimesta kehitetty analytiikkatyökalu, sen tärkeimmät ominaisuudet ja toiminnot. Analytiikkatyökalun pilottiversion kehitystyö noudatti verkkoanalytiikan viitekehystä ja se auttoi myös opinnäytetyön empiirisen osuuden kuvaamista loogisessa järjestyksessä. Haasteellista empiirisen osuuden kirjoitustyön kannalta oli, että dokumentaatiota ei juurikaan pilotista demon lisäksi syntynyt. Opinnäytetyön case –esimerkin kirjoitustyö oli lähinnä omien muistiinpanojeni varas-

sa. Empiirisen osuuden kirjoittamisen vaikeudeksi osoittautui asian esittäminen niin ytimekkäästi ja selkokielisesti, että analytiikkatyökalun toiminnollisuudesta saa käsityksen ilman, että lukija näkee demoa.

Opinnäytetyön kirjoittaminen oli aikaa vievää ja työlästä, mutta kehittävää ja palkitsevaa. Lähdemateriaalia aiheen ympäriltä löytyy lukuisia etenkin englanninkielistä kirjallisuutta. Kuitenkin oikeanlaisen ja ajantasaisen materiaalin löytäminen tuotti aika-ajoin haasteita. Tiedonhaku ja materiaaliin perehtyminen oli ehdottomasti opinnäytetyöprosessin työläin osuus. Työn etenemistä: tiedon hakua ja sen jäsentämistä helpotti tietovarastoinnin, liiketoimintatiedon raportoinnin ja analysoinnin alueelta työni kautta kertynyt kokemus. Opinnäytetyön suunnittelun aloitin keväällä 2014. Tavoitteena oli saada työ valmiiksi viimeistään kevään 2015 aikana. Empiirisen osuuden kirjoittamiseen vaikutti Need For Speed –ohjelman Invencon osuuden edistyminen ja valmistuminen. Tästä johtuen opinnäytetyön valmistuminen venyi syksyyn. Näin jälkiviisaana opinnäytetyön suunnittelun ja varsinaisen kirjoittamisen olisi voinut aloittaa myöhemmin. Nyt prosessin läpivienti tuntui ikuisuusprojektilta.

Opinnäytetyö mahdollisti uusiin asioihin perehtymisen ja tietämyksen täydentämisen ennestään tutuilta liiketoiminnan hallinnan alueilta. Vaikka teen työtä päivittäin tiedon ja tiedon jalostamisen kanssa, syventyi tietämykseni big datasta, verkkoanalytiikasta ja tiedolla johtamisesta. Erittäin mielenkiintoiseksi koin analytiikkatyökalun pilottiversion työstämisen. Teoria muuttui käytännöksi Need For Speed –ohjelman myötä. Konkreettisesti näin miten mittarit suunnitellaan ja toteutetaan asiakkaan tarpeet huomioiden.

Lähteet

Big data 2015. Big data – määritelmiä. Luettavissa: <http://www.bigdata.fi/big-data-maaritelma>. Luettu: 16.6.2015.

Cheesman, P. 2014. How to Describe Big Data? Luettavissa: <http://www.patrickcheesman.com/how-big-data-can-transform-your-understanding-of-your-customers/>. Luettu: 26.7.2015.

Clay, B. 2015. Web Analytics. Luettavissa: <http://www.bruceclay.com/eu/analytics/google-analytics.htm>. Luettu: 21.7.2015.

Data Science Series 2012. Ten Practical Big Data Benefits. Luettavissa: <http://datascienceseries.com/stories/ten-practical-big-data-benefits>. Luettu: 14.7.2015.

Digile 2015. N4S-ohjelma: suomalaiset ohjelmistoyritykset nopeuttavat digitaalista taloutta. Luettavissa: <http://www.n4s.fi/fi/>. Luettu: 28.7.2015.

Digital Analytics Association 2015. What is Digital Analytics? . Luettavissa: <http://www.digitalanalyticsassociation.org/>. Luettu: 17.6.2015.

Digital Insights 2014. Social Media Statistics 2014. Luettavissa: <http://blog.digitalinsights.in/social-media-users-2014-stats-numbers/05205287.html>. Luettu: 18.5.2015.

Fair Isaac Corporation 2015. When is Big data the Way to Customer Centricity? Luettavissa: http://www.fico.com/landing/pdf/67_Big_Data_Customer_Centricity_2951WP.pdf. Luettu: 17.6.2015.

Fonecta 2015. Miten hyödynnät online-analytiikkaa? Luettavissa: <https://hyotytieto.fonecta.fi/hyotytieto/viikon-hyotytieto-miten-hyodynnat-online-analytiikkaa>. Luettu: 21.7.2015.

Garside, W., Cox, B. 2013. Big Data Storage for Dummies. Wiley. West Sussex.

IT Business Edge 2015a. Structured data. Luettavissa: http://www.webopedia.com/TERM/S/structured_data.html. Luettu: 18.8.2015.

IT Business Edge 2015b. Unstructured data. Luettavissa:

http://www.webopedia.com/TERM/U/unstructured_data.html. Luettu:18.8.2015.

Jyväskylän yliopisto 2015a. Laadullinen tutkimus. Luettavissa:

<https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/tutkimusstrategiat/laadullinen-tutkimus>. Luettu:20.8.2015

Jyväskylän yliopisto 2015b. Määrällinen tutkimus. Luettavissa:

<https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/tutkimusstrategiat/määrällinen-tutkimus>. Luettu: 20.8.2015

Kaushik, A. 2010. Web Analytics 2.0. Wiley. Indianapolis.

Krishnan, K. 2013. Data Warehousing in the Age of Big Data. Luettavissa:

http://proquest.safaribooksonline.com/book/databases/data-warehouses/9780124058910/part-1-big-data/chp002_html. Luettu: 21.7.2015.

Liikenne- ja viestintävirasto 2014. Big datan hyödyntäminen. Luettavissa:

http://www.lvm.fi/c/document_library/get_file?folderId=3082174&name=DLFE-24783.pdf&title=Julkaisuja+20-2014. Luettu:14.7.2015.

Lukion filosofia 2001. Tiedon klassinen määritelmä. Luettavissa:

http://www.mv.helsinki.fi/home/olappi/lukionfilosofia/filosofit/platon/klassinen_tiedon_maaritelma.htm. Luettu: 18.8.2015.

Microsoft 2015. Reveal new insights with Big Data. Luettavissa:

<http://www.microsoft.com/en-us/server-cloud/solutions/big-data.aspx>. Luettu: 14.7.2015.

Microsoft Azure 2015. HDInsight. Luettavissa: <http://azure.microsoft.com/en-us/services/hdinsight/>. Luettu: 14.7.2015.

Netcraft 2015. April 2015 Web Server survey. Luettavissa:

<http://news.netcraft.com/archives/category/web-server-survey/>. Luettu: 18.5.2015.

Mohanty, S., Jagadeesh, M., Srivatsa, H. 2013. Big Data Imperatives. APress. New York.

Prajapati, V. 2013. Big Data Analytics with R and Hadoop. Luettavissa:
http://proquest.safaribooksonline.com/book/databases/hadoop/9781782163282/preface/pr07s04_html#X2ludGVybmFsX0h0bWxWaWV3P3htbGlkPTk3ODE3ODIxNjMyODIIMkZwcjA3czA0X2h0bWwmcXVlcnk9. Luettu: 21.7.2015.

Tietojohtamisen tutkimuskeskus NOVI 2014. Tietojohtamisen perusteet: tiedon tasot ja lajit. Luettavissa: <http://www.slideshare.net/Noviresearch/tijop-osa3-tiedontasot>. Luettu: 16.6.2015.

Turner, V., Gantz, J., Reinsel, D., Minton, S. 2014. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. Luettavissa:
<http://idcdocserv.com/1678>. Luettu: 15.7.2015.

Vakkuri, M. 20.6.2013. Big data muuttaa maailmaa. Talouselämä kumppaniblogi. Luettavissa:
<http://www.talouselama.fi/kumppaniblogit/tieto/big+data+muuttaa+maailmaa/a2191461>. Luettu: 18.8.2015.

Web Analytics Association 2008. Web Analytics Definition. Luettavissa:
http://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAnalyticsDefinitions.pdf. Luettu: 17.6.2015.